

19 **BUNDESREPUBLIK
DEUTSCHLAND**



**DEUTSCHES
PATENTAMT**

12 **Übersetzung der
europäischen Patentschrift**
87 **EP 0 561 752 B 1**
10 **DE 693 18 209 T 2**

7A-80942
51 Int. Cl. 6:
G 10 L 5/04
G 10 L 5/00
G 10 L 3/00
G 10 L 7/06

- 21 Deutsches Aktenzeichen: 693 18 209.1
86 Europäisches Aktenzeichen: 93 850 026.1
86 Europäischer Anmeldetag: 8. 2. 93
87 Erstveröffentlichung durch das EPA: 22. 9. 93
87 Veröffentlichungstag
der Patenterteilung beim EPA: 29. 4. 98
87 Veröffentlichungstag im Patentblatt: 27. 8. 98

30 Unionspriorität:
9200817 17. 03. 92 SE

73 Patentinhaber:
Televerket, Farsta, SE

74 Vertreter:
Glawe, Delfs, Moll & Partner, Patentanwälte, 20148
Hamburg

84 Benannte Vertragsstaaten:
BE, CH, DE, FR, LI, NL

72 Erfinder:
Kaja, Jaan, S-136 80 Haninge, SE

54 Verfahren und Anordnung zur Sprachsynthese

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patentamt inhaltlich nicht geprüft.

DE 693 18 209 T 2

11.12.97

EP 93 850 026.1
TELEVERKET
Scho/N/he (he)

5

Die vorliegende Erfindung betrifft ein Verfahren und eine An-
ordnung zur Sprachsynthese und stellt einen automatischen Me-
chanismus zum Simulieren von menschlicher Sprache bereit. Das
10 erfindungsgemäße Verfahren stellt eine Anzahl von Steuerpara-
metern zum Steuern einer Sprachsynthesevorrichtung bereit.

In natürlicher Sprache überlappen sich die darin enthaltenen
15 Phoneme miteinander. Diese Erscheinung wird als gleichzeitige
Artikulation bezeichnet. Wie nachfolgend dargelegt wird,
verwenden das erfindungsgemäße Sprachsynthese-Verfahren und
die erfindungsgemäße Sprachsynthese-Anordnung eine zweistim-
mige (diphonische) Synthese zum Erzeugen von Sprache mittels
20 Formantensynthese. Ein Interpolationsmechanismus handhabt au-
tomatisch die gleichzeitige Artikulation. Überdies stellt
die vorliegende Erfindung die Möglichkeit zur mehrstimmigen
(polyphonen) Synthese, insbesondere zur zweistimmigen
(diphonischen) Synthese, aber auch zur triphonen Synthese
25 und quadrophonen Synthese bereit.

Es ist bekannt, daß die Synthese von Text und/oder Sprache
häufig mit einer syntaktischen Analyse des Textes beginnt, bei
der Worten, die auf mehr als eine Art interpretiert werden
30 können, eine korrekte Aussprache gegeben wird, das heißt, es
wird eine geeignete phonetische Übertragung gewählt. Ein Bei-
spiel davon ist das schwedische Wort "buren", das als ein
Substantiv oder als die Partizipform eines Verbs interpre-
tiert werden kann.

35

Durch Verwendung der syntaktischen Analyse und der Silben-
struktur eines Satzes als einen Anfangspunkt kann eine grund-
legende Tonkurve für die ganze Phrase gebildet werden, und
die Dauer der darin enthaltenen Phoneme kann bestimmt werden.

Nach diesem Prozeß können die Phoneme in einer Anzahl verschiedener Arten akustisch realisiert werden.

5 Ein bekanntes Verfahren zur Sprachsynthese ist die Formanten-
synthese. Mit diesem Verfahren wird die Sprache durch Anwen-
den verschiedener Filter auf eine Quelle erzeugt. Die Filter
werden mittels einer Anzahl von Steuerparametern gesteuert,
die u.a. Formanten, Bandbreiten und Quellenparameter ein-
schließen. Ein Prototypsatz an Steuerparametern wird durch
10 ein Allophon gespeichert. Eine gleichzeitige Artikulierung
wird durch Bewegen von Start/Endpunkten der Steuerparameter
mit der Hilfe von Regeln, d.h. Regelsynthese gehandhabt. Ein
Problem mit diesem Verfahren ist es, daß es eine große Menge
von Regeln zum Handhaben der vielen möglichen Kombinationen
15 von Phonemen benötigt. Überdies ist das Verfahren schwierig
zu überwachen.

Ein weiteres bekanntes Verfahren zur Sprachsynthese ist die
zweistimmige (diphonische) Synthese. Mit diesem Verfahren
20 wird die Sprache durch Verbinden von Segmenten aus aufgenom-
menen Wellenformen von aufgenommener Sprache erzeugt, und die
erwünschte Basistonkurve und Dauer wird durch Signalverarbei-
tung erzeugt. Eine zugrundeliegende Voraussetzung dieses Ver-
fahrens ist es, daß es in jedem zweistimmigen Laut einen Be-
25 reich gibt, der spektral gleichbleibend ist, und daß dort
spektrale Ähnlichkeit vorherrscht; andernfalls wird dort eine
spektrale Diskontinuität erhalten, die ein Problem ist. Es
ist mit diesem Verfahren auch schwierig, die Wellenformen
nach dem Aufnehmen und Segmentation zu ändern. Es ist eben-
30 falls schwierig, Regeln anzuwenden, da die Wellenformsegmente
fest sind.

Es gibt keine Probleme mit spektralen Diskontinuitäten in der
Formantensprachsynthese. Die zweistimmige (diphonische)
35 Sprachsynthese benötigt keine Regeln zum Handhaben des Pro-
blems der gleichzeitigen Artikulierung.

WO-A-90/13890 offenbart ein Verfahren und eine Vorrichtung zum Codieren einer elektronischen Wellenform als ein digitales Signal und insbesondere die Codierung und Erzeugung von Tonsignalen, besonders jenen, die Sprache einschließen. Gemäß diesem Verfahren werden Werte von alternativen Maxima und Minima in der Wellenform, zum Beispiel einem Tonsignal, aus der Wellenform extrahiert und mit verbundener zeitlicher Information kombiniert und als das Digitalsignal angeordnet. Die Wellenform wird aus dem Digitalsignal regeneriert, indem Segmente einer vorherbestimmten Wellenfunktion, zum Beispiel einer Cosinuswelle einer Periode, die durch die zeitliche Information bestimmt wird, und einer Amplitude, die durch die Werte der Maxima und Minima bestimmt, miteinander verbunden werden.

Es ist ein Ziel der vorliegenden Erfindung, ein zweistimmiges (diphonisches) Syntheseverfahren zu verwenden, das heißt die Verwendung gespeicherter Steuerparameter, die durch Kopieren natürliche Sprache mit der Hilfe von Synthese extrahiert worden sind, zum Erzeugen von Sprache mittels Formantensynthese. Ein Interpolationsmechanismus handhabt automatisch die gleichzeitige Artikulierung. Wenn es nichtsdestoweniger wünschenswert ist, Regeln anzuwenden, dann kann dies in der Tat geschehen.

Die Erfindung stellt ein Verfahren zur Sprachsynthese bereit, wobei Steuerparameter bestimmt werden, die u.a. Formanten, Bandbreiten und Quellenparameter einschließen, die zum Steuern der Synthese von Sprache erforderlich sind, und wobei die Steuerparameter in einer Matrix oder Sequenzliste für jeden mehrstimmigen Laut gespeichert werden, dadurch gekennzeichnet, daß das Verfahren eine zweistimmige (diphonische) Synthese zum Erzeugen synthetischer Sprache mittels Formantensynthese und einen Interpolationsmechanismus zum automatischen Handhaben gleichzeitiger Artikulierung verwendet, und daß das Verfahren die Schritte des Definierens des Verhaltens der entsprechenden Steuerparameter hinsichtlich der Zeit um jede Phonemgrenze herum und das Verbinden der mehrstimmigen

Laute durch Bilden eines gewichteten Mittelwerts der Kurven einschließt, die durch ihre entsprechenden gespeicherten Steuerparameter definiert werden. Die Bildung der Steuerparameter kann durch numerische Analyse bewirkt werden, die die Simulation natürlicher Sprache einbezieht.

Die Dauer des Phonems, das in dem entsprechenden mehrstimmigen Laut enthalten ist, kann an den benachbarten mehrstimmigen Laut durch Quantisieren der Dauer für ein Parameter-Abtastintervall angepaßt werden. Der gewichtete Mittelwert kann durch Multiplikation mit einer Gewichtsfunktion, wie zum Beispiel einer Cosinusfunktion gebildet werden.

In einem bevorzugten Verfahren der vorliegenden Erfindung sind die mehrstimmigen Laute zweistimmige Laute, wobei jeder der zweistimmigen Laute erste und zweite Phoneme aufweist und das Verfahren die Schritte einschließt: Speichern eines Satzes zweistimmiger Laute auf der Grundlage von Formantensynthese; Definieren einer Kurve für jeden Steuerparameter, wobei die Kurve das Verhalten der Parameter mit der Zeit um die Phonemgrenze herum beschreibt; und Verbinden zweier zweistimmiger Laute miteinander durch Bilden eines gewichteten Mittelwerts zwischen dem zweiten Phonem in einem der zweistimmigen Laute und dem ersten Phonem im anderen der zweistimmigen Laute. Die Kurve kann für einen zweiten Formanten für die beiden zweistimmigen Laute definiert werden, wobei in diesem Fall der eine der zweistimmigen Laute einen ersten Teil oder Anfang eines Tons darstellt und der andere der zweistimmigen Laute einen zweiten Teil oder Ende des Tons darstellt, wobei der Ton durch Verbinden der ersten und zweiten Teile miteinander gebildet wird.

Die Erfindung stellt auch eine Anordnung zum Bilden synthetisch Tonkombinationen bereit, die ein erfindungsgemäßes Verfahren verwendet, wie es in den vorhergehenden Abschnitten dargelegt wird.

Die Erfindung stellt ferner eine Anordnung zum Bilden synthetisch Tonkombinationen bereit, die Mittel zum Bestimmen von Steuerparametern, die u.a. Formanten, Bandbreiten und Quellenparameter einschließen, die zum Steuern der Bildung von synthetischen Tonkombinationen erforderlich sind und Steuerparameter-Speichermittel für jeden mehrstimmigen Laut einschließen, dadurch gekennzeichnet, daß die Anordnung eine zweistimmige (diphonische) Synthese zum Erzeugen synthetischer Sprache mittels Formantensynthese und einen Interpolationsmechanismus zum automatischen Handhaben gleichzeitiger Artikulierung verwendet, und daß die Anordnung Mittel zum Definieren des Verhaltens der entsprechenden Steuerparameter hinsichtlich der Zeit um jede Phonemgrenze herum und zum Verbinden der mehrstimmigen Laute durch Bilden eines gewichteten Mittelwerts der Kurven einschließt, die durch ihre entsprechenden gespeicherten Steuerparameter definiert werden.

Gemäß der erfindungsgemäßen Anordnung kann die Dauer des Phonems, das in dem entsprechenden mehrstimmigen Laut enthalten ist, an den benachbarten mehrstimmigen Laut durch Quantisieren der Dauer für ein Parameter-Abtastintervall angepaßt werden, und der gewichtete Mittelwert kann durch Multiplikation mit einer Gewichtsfunktion, wie zum Beispiel einer Cosinusfunktion gebildet werden. Die Anordnung kann Mittel zum numerischen Analysieren zum Bilden der Steuerparameter einschließen.

In einer bevorzugten erfindungsgemäßen Anordnung, wobei die mehrstimmigen Laute zweistimmige Laute sind, wobei jeder zweistimmige Laut erste und zweite Phoneme aufweist, können die Speichermittel angepaßt sein, einen Satz zweistimmiger Laute auf der Grundlage von Formantensynthese zu speichern, und die das Verhalten definierenden Mittel können angepaßt sein, eine Kurve für jeden Steuerparameter zu definieren, wobei jede der Kurven das Verhalten eines entsprechenden Parameters mit der Zeit um die Phonemgrenze herum beschreibt, wobei die beiden zweistimmigen Laute durch Bilden eines gewichteten Mittelwerts zwischen dem zweiten Phonem in einem der

zweistimmigen Laute und dem ersten Phonem im anderen der zweistimmigen Laute miteinander verbunden werden. Die Kurve kann für einen zweiten Formanten für die beiden zweistimmigen Laute definiert werden, wobei einer der zweistimmigen Laute einen ersten Teil oder Anfang eines Tons darstellt und der andere der zweistimmigen Laute einen zweiten Teil oder Ende des Tons darstellt, wobei der Ton durch Verbinden der ersten und zweiten Teile miteinander gebildet wird.

Die vorhergehenden und anderen erfindungsgemäßen Merkmale werden aus der folgenden Beschreibung unter Bezugnahme auf die einzige Figur der beigefügten Zeichnungen besser zu verstehen sein, die ein Diagramm ist, das das Verbinden der zweistimmigen Laute gemäß der vorliegenden Erfindung zeigt.

Natürliche menschliche Sprache kann in Phoneme unterteilt werden. Ein Phonem ist die kleinste Komponente mit semantischen Unterschied der Sprache. Ein Phonem kann als solches durch verschiedene Töne, Allophone, realisiert werden. In der Sprachsynthese muß festgestellt werden, welches Allophon für ein bestimmtes Phonem verwendet werden sollte, jedoch ist dies kein Gegenstand für die vorliegende Erfindung.

Es gibt eine Kopplung zwischen den unterschiedlichen Teilen im Sprachorgan, zum Beispiel zwischen der Zunge und dem Kehlkopf, und die Artikulatoren, Zunge, Mund und so weiter, können nicht sofort von einer Stelle zu einer anderen bewegt werden. Es gibt daher eine starke gleichzeitige Artikulierung zwischen den Phonemen; folglich beeinflussen die Phoneme einander. Um Sprache, die naturgetreu ist, aus einer Sprachsynthesevorrichtung zu erhalten, muß sie daher fähig sein, gleichzeitige Artikulierung zu handhaben.

Die vorliegende Erfindung sorgt auch für eine Sprachsynthese mehrstimmiger Laute, das heißt die Verbindung von mehreren Phonemen, zum Beispiel Triphon-Synthese oder Quadrophon-Synthese. Dies kann wirksam mit bestimmten Vokaltönen benutzt werden, die keine gleichbleibenden Teile aufweisen, die zum

- Verbinden geeignet sind. Bestimmte Kombinationen von Konsonanten sind auch störend. In natürlicher menschlicher Sprache gibt es immer irgendwo Bewegung, und der nächste Ton wird vorweggenommen. Zum Beispiel wird in dem Wort "sprite" das Sprachorgan für den Vokal geformt, bevor das "s" ausgesprochen wird. Durch Speichern des Triphons als Punkte längs einer Kurve kann das Triphon mit dem nachfolgenden Phonem verbunden werden.
- Die Wellenform der Sprache kann mit der Reaktion aus einer Resonanzkammer, der Stimmröhre, auf eine Reihe von Impulsen, quasiperiodische Stimmbandimpulse in stimmhaftem Ton, oder Tönen, die mit einer Zusammenschnürung in stimmlosen Tönen erzeugt werden, verglichen werden. Bei der Sprachvorhersage bildet die Stimmröhre einen akustischen Filter, in dem Resonanz in der verschiedenen Hohlräumen auftritt, die in diesem Zusammenhang gebildet werden. Die Resonanzen werden als Formanten bezeichnet, und sie erscheinen im Spektrum als Energiespitzen bei den Resonanzfrequenzen. Bei kontinuierlicher Sprache variieren die Formantenfrequenzen mit der Zeit, da die Resonanzhohlräume ihre Stellung verändern. Die Formanten sind daher zum Beschreiben des Tons von Wichtigkeit und können zum Steuern von Sprachsynthese verwendet werden. Eine Sprachphrase wird mit einer geeigneten Aufnahmeanordnung aufgenommen und in einem Medium gespeichert, das zur Datenverarbeitung geeignet ist. Die Sprachphrase wird analysiert, und geeignete Steuerparameter werden gemäß einer der unten dargelegten Verfahren gespeichert.
- Die Speicherung der oben erwähnten Steuerparameter kann durch irgendeines der folgenden Verfahren bewirkt werden:
- (1) Eine Matrix wird gebildet, in der jeder Reihenvektor einem Parameter entspricht und die Elemente darin den abgetasteten Parameterwerten entsprechen. (Eine typische Abtastfrequenz ist 200 Hz). Dieses Verfahren ist für eine Synthese zweistimmiger Laute geeignet.

- (2) Eine Sequenz mathematischer Funktionen, Start/Endwerte + Funktion, wird für jeden Parameter gebildet. Dieses Verfahren ist für eine Synthese mehrstimmiger Laute geeignet und macht es möglich, wenn gewünscht, Regeln der gebräuchlichen Art zu verwenden.

Ein Verfahren zum Erzeugen gespeicherter Steuerparameter, das eine gute Synthesequalität liefert, ist es, ein Kopieren einer Synthese einer natürlichen Phrase auszuführen. Mit dieser Anordnung werden numerische Verfahren in einem iterativen Prozeß verwendet, der in Schritten sicherstellt, daß die synthetische Phrase mehr und mehr der natürlichen Phrase ähnelt. Wenn eine ausreichend gute Ähnlichkeit erhalten worden ist, können die Steuerparameter, die dem gewünschten zweistimmigen Laut/mehrstimmigen Laut entsprechen, aus der synthetischen Phrase extrahiert werden.

Die vorliegende Erfindung löst das Problem der gleichzeitigen Artikulierung durch Verwendung eines Interpolationsverfahrens. So wird ein Satz von zweistimmigen Lauten auf der Grundlage von Formantensynthese gespeichert. Für jeden Parameter wird eine Kurve gemäß entweder dem Verfahren (1) oder Verfahren (2) wie oben dargelegt definiert, die das Verhalten des Parameters mit Zeit um die Phonemgrenze herum beschreibt.

Zwei zweistimmige Laute werden durch Bilden eines gewichteten Mittelwerts zwischen dem zweiten Phonem im ersten zweistimmigen Laut und dem ersten Phonem im zweiten zweistimmigen Laut miteinander verbunden.

Die einzige Figur der beigefügten Zeichnungen zeigt den erfindungsgemäßen Verbindungsmechanismus im Detail. Die Kurven stellen einen Parameter, zum Beispiel den zweiten Formanten für die beiden zweistimmigen Laute, dar. Der erste zweistimmige Laut kann zum Beispiel der Ton "ba" sein, und der zweite zweistimmige Laut kann der Ton "ad" sein, die, wenn sie miteinander verbunden werden, zu "bad" werden. Die Kurven gehen

nach links und rechts asymptotisch zu konstanten Werten weiter.

5 Im mittleren Phonem ist ein Interpolationsmechanismus in Betrieb. Die beiden Kurven der zweistimmigen Laute werden jeweils mit ihrer eigenen Gewichtsfunktion gewichtet, die im unteren Teil der einzigen Figur der beigefügten Zeichnungen
10 gezeigt wird. Die Gewichtsfunktionen sind vorzugsweise Cosinusfunktionen, um einen glatten Übergang zu erhalten, jedoch ist dies nicht entscheidend, da auch lineare Funktionen verwendet werden können.

Bestimmte Gebiete werden nicht interpoliert, da bestimmte
15 Sprachtöne, wie Stoppkonsonanten, einen Druck mit einbeziehen, der im Mundhohlraum aufgebaut wird, der dann freigegeben wird, zum Beispiel "pa". Der Prozeß von der Zeit, zu der der Druck freigegeben wird, bis die Stimmbandimpulse erzeugt werden, ist rein mechanisch und wird nicht merklich von der verbleibenden Länge des Phonems in der Phrase beeinflusst. Sollte
20 die Dauer des Stoppkonsonanten ausgedehnt sein, ist es die stumme Phrase, die länger wird. Der Interpolationsmechanismus muß es daher vermeiden, Bestimmte Bits auszudehnen. Um die Segmentgrenzen herum ist es daher für bestimmte Bits notwendig, eine feste Länge aufzuweisen, das heißt, die Anwendung
25 der Gewichtsfunktionen beginnt ein Bit nach der Segmentgrenze und endet ein Bit vor der Segmentgrenze.

Es ist die syntaktische Analyse, die bestimmt, wie eine Phrase synthetisiert wird. Unter anderem werden die grundlegende
30 Tonkurve und Dauer der Segmente bestimmt, was unter anderem eine unterschiedliche Betonung liefert. Die Betonung wird zum Beispiel durch Ausstrecken des Segmentes und einer Krümmung in der grundlegenden Tonkurve erzeugt, während die Amplitude geringere Bedeutung aufweist.

35 Die Segmente können eine unterschiedliche Dauer aufweisen, das heißt Zeitlänge. Die Segmentgrenzen werden durch den Übergang von einem Phonem zum nächsten Phonem bestimmt, wäh-

rend die syntaktische Analyse bestimmt, wie lange ein Phonem sein soll. Jedes Phonem weist einen ästhetischen Wert auf. Die Kurven oder die Funktionen können zum Anpassen zweier Dauern aneinander gestreckt werden. Dies geschieht durch
5 Quantisieren für ein ms-Intervall und Manipulieren der Kurven. Dies wird ebenfalls durch die Kurven erleichtert, die gegen unendlich asymptotisch sind.

Das erfindungsgemäße Verfahren stellt Steuerparameter bereit,
10 die direkt in einem herkömmlichen Sprachsynthesystem verwendet werden können. Die vorliegende Erfindung stellt auch eine Anordnung zur Sprachsynthese, d.h. Bildung synthetischer Tonkombinationen innerhalb ausgewählter Zeitintervalle bereit. Durch Verwendung einer zweistimmigen (diphonischen)
15 Sprachsynthesetechnik, d.h. die Verwendung gespeicherter Steuerparameter, die durch Kopieren natürlicher Sprache mit der Hilfe von Synthese extrahiert worden sind, um Sprache mittels Formantensynthese zu erzeugen, kann eine naturgetreue Sprache erhalten werden, da die Formantensynthese
20 glatte Kurven liefert, die ohne irgendwelche Diskontinuitäten verbunden werden.

EP 93 850 026.1
Televerket

5

Patentansprüche

1. Verfahren zur Sprachsynthese, bei dem Steuerparameter einschließlich u.a. Formanten, Bandbreiten und Quellenparameter bestimmt werden, die zum Steuern der Sprachsynthese erforderlich sind, und bei dem die Steuerparameter in einer Matrix oder Sequenzliste für jeden mehrstimmigen Laut gespeichert werden, dadurch gekennzeichnet, daß das Verfahren zweistimmige (diphonische) Synthese zum Erzeugen synthetischer Sprache mit Hilfe von Formantensynthese und einen Interpolationsmechanismus zum automatischen Handhaben gleichzeitiger Artikulierung verwendet, und daß das Verfahren die Schritte einschließt, das Verhalten der entsprechenden Steuerparameter in bezug auf die Zeit um jede Phonemgrenze herum zu definieren und die mehrstimmigen Laute durch Bilden eines gewichteten Mittelwertes der Kurven zu verbinden, die durch ihre entsprechenden gespeicherten Steuerparameter definiert sind.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Dauer des Phonems, das in dem entsprechenden mehrstimmigen Laut enthalten ist, an den benachbarten mehrstimmigen Laut angepaßt wird, indem die Dauer für ein Parameter-Abtastintervall quantisiert wird.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß der gewichtete Mittelwert durch Multiplikation mit einer Gewichtsfunktion gebildet wird.
4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß der gewichtete Mittelwert durch Multiplikation mit einer Cosinusfunktion gebildet wird.
5. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die Bildung der Steuerparameter durch

numerische Analyse einschließlich Simulation von natürlicher Sprache bewirkt wird.

- 5 6. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die vielstimmigen Laute zweistimmige Laute sind, wobei jeder zweistimmige Laut erste und zweite Phoneme aufweist, und daß das Verfahren die Schritte einschließt, einen Satz von zweistimmigen Lauten aufgrund von Formantensynthese zu speichern, eine Kurve für jeden Steuerparameter zu definieren, welche Kurve das Verhalten des Parameters mit der Zeit um die Phonemgrenze herum beschreibt, und zwei zweistimmige Laute miteinander zu verbinden, indem ein gewichteter Mittelwert zwischen dem
10 zweiten Phonem in einem der zweistimmigen Laute und dem ersten Phonem in dem anderen der Laute gebildet wird.
15
7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß die Kurve für einen zweiten Formanten der beiden zweistimmigen Laute definiert wird, daß einer der zweistimmigen Laute einen ersten Teil oder Anfang eines Tons und der andere zweistimmige Laut einen zweiten Teil oder Ende des Tons darstellt, und daß der Ton durch Verbinden der ersten und zweiten Teile miteinander gebildet wird.
20
- 25 8. Anordnung zum Bilden von synthetischen Tonkombinationen unter Verwendung eines Verfahrens nach einem der vorangehenden Ansprüche.
- 30 9. Anordnung zum Bilden synthetischer Tonkombinationen, die Mittel zum Bestimmen von Steuerparametern einschließlich u.a. Formanten, Bandbreiten und Quellenparametern, die zum Steuern der Bildung von synthetischen Tonkombinationen erforderlich sind, und Steuerparameter-Speichermittel für jeden mehrstimmigen Laut einschließt, dadurch gekennzeichnet, daß die Anordnung zweistimmige (diphonische)
35 Synthese zum Erzeugen von synthetischer Sprache mit Hilfe von Formantensynthese und einen Interpolationsmechanismus zum automatischen Handhaben der gleichzeitigen Artikula-

- tion verwendet, und daß die Anordnung Mittel zum Definieren des Verhaltens der Steuerparameter in bezug auf die Zeit um jede Phonemgrenze herum und zum Verbinden der vielstimmigen Laute durch Bilden eines gewichteten Mittelwertes der Kurven einschließt, die durch ihre gespeicherten Steuerparameter definiert sind.
- 5
10. Anordnung nach Anspruch 9, dadurch gekennzeichnet, daß die Dauer des Phonems, das in dem entsprechenden mehrstimmigen Laut enthalten ist, an den benachbarten mehrstimmigen Laut angepaßt wird, indem die Dauer für ein Parameter-Abtastintervall quantisiert wird.
- 10
11. Anordnung nach Anspruch 9 oder 10, dadurch gekennzeichnet, daß der gewichtete Mittelwert durch Multiplikation mit einer Gewichtsfunktion gebildet wird.
- 15
12. Anordnung nach Anspruch 11, dadurch gekennzeichnet, daß der gewichtete Mittelwert durch Multiplikation mit einer Cosinusfunktion gebildet wird.
- 20
13. Anordnung nach einem der Ansprüche 9 bis 12, dadurch gekennzeichnet, daß die Anordnung Mittel zum numerischen Analysieren zum Bilden der Steuerparameter einschließt.
- 25
14. Anordnung nach einem der Ansprüche 9 bis 13, dadurch gekennzeichnet, daß die mehrstimmigen Laute zweistimmige Laute sind, wobei jeder zweistimmige Laut erste und zweite Phoneme aufweist, daß die Speichermittel zum Speichern eines Satzes von zweistimmigen Lauten aufgrund von Formantensynthese ausgebildet sind, und daß die das Verhalten definierenden Mittel zum Definieren einer Kurve für jeden Steuerparameter ausgebildet sind, wobei jede der Kurven das Verhalten eines entsprechenden Parameters mit der Zeit um die Phonemgrenze herum beschreibt, wobei die beiden zweistimmigen Laute durch Bilden eines gewichteten Mittelwertes zwischen dem zweiten Phonem in einem der zweistimmigen Laute und dem ersten Phonem in dem an-
- 30
- 35

deren der zweistimmigen Laute miteinander verbunden werden.

- 5 15. Anordnung nach Anspruch 14, dadurch gekennzeichnet, daß die Kurve für einen zweiten Formanten der beiden zweistimmigen Laute definiert wird, daß der eine der zweistimmigen Laute einen ersten Teil oder Anfang eines Tons und der andere der zweistimmigen Laute einen zweiten Teil oder Ende des Tons darstellt, und daß der Ton durch Verbinden der ersten und zweiten Teile gebildet wird.
- 10

11.12.97

EP 93 850 026.1

1/1

